DOCUMENT RESUME

ED 416 677 FL 025 072

AUTHOR	Rhodes, Nancy C.
TITLE	Alternative Assessment for Immersion Students: The Student Oral Proficiency Assessment (SOPA).
INSTITUTION	Center for Applied Linguistics, Washington, DC.
PUB DATE	1996-00-00
NOTE	22p.; Paper presented at the European Conference on Immersion Programmes (3rd, Barcelona, Spain, September 25-28, 1996).
PUB TYPE	Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Elementary Education; Evaluation Criteria; Evaluation Methods; *Immersion Programs; Interviews; Language Fluency; *Language Tests; Listening Comprehension; Rating Scales; Second Language Instruction; *Spanish; Speech Skills; *Student Evaluation; Test Construction; Test Use; Testing; *Verbal Tests; Young Children
IDENTIFIERS	ACTFL Proficiency Guidelines; Alternative Assessment; *Oral Proficiency Testing; *Student Oral Proficiency Assessment; Two Way Bilingual Education

ABSTRACT

An alternative assessment instrument developed to assess oral language skills of students in Spanish immersion programs in the United States is described. Originally developed by the Center for Applied Linguistic to evaluate 6-year-old immersion students' speaking and listening skills in a school in Oak Ridge, Tennessee, the Student Oral Proficiency Assessment (SOPA) is now used as a prototype for oral language assessment of 6-to-9-year-olds students in a variety of immersion program types. The SOPA interview consists of four parts: listening comprehension; informal questions; science and language usage; and story telling. Two students are assessed at a time by two examiners in a non-stressful, friendly environment. The goal of the assessment is to show what the students can do with language, not what they cannot do. Students' comprehension and fluency are rated on a six-level holistic scale based on a modified rating scale of the American Council on the Teaching of Foreign Languages (ACTFL). In addition to a description of the instrument and rating scale, results from a two-way Spanish immersion program are presented, and plans for collection reliability and validity data are discussed. (Author/MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 416 677

Alternative Assessment for Immersion Students: The Student Oral Proficiency Assessment (SOPA)

Nancy C. Rhodes
Center for Applied Linguistics
Washington, DC USA

CAL

Submitted for the Proceedings of the III European Conference on Immersion Programmes
Barcelona, Spain, September 25-28, 1996

FL025072

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Nancy Rhodes
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

BEST COPY AVAILABLE

Alternative Assessment for Immersion Students: The Student Oral Proficiency Assessment (SOPA)

Nancy C. Rhodes
Center for Applied Linguistics
Washington, DC USA

Abstract

This paper describes an alternative assessment instrument that has been developed to assess oral language skills of students in Spanish immersion programs in the United States. Originally developed by the Center for Applied Linguistics to evaluate six-year-old immersion students' speaking and listening skills in a school in Oak Ridge, Tennessee, the Student Oral Proficiency Assessment (SOPA) is now being used as a prototype for oral language assessment of six-to-nine-year-old students in a variety of types of immersion programs. The SOPA interview consists of four parts: listening comprehension, informal questions, science and language usage, and story telling. Two students are assessed at a time by two examiners in a non-stressful, friendly environment. The goal of the assessment is to show what the students *can* do with language, not what they cannot do. Students' comprehension and fluency is rated on a six-level holistic scale based on a modified rating scale of the American Council on the Teaching of Foreign Languages. In addition to describing the instrument and rating scale, results from a two-way Spanish immersion program will be presented, and plans for collecting reliability and validity data will be discussed.

I. Introduction

With the dramatic increase in the number of language immersion programs around the world in the last two decades, there has been increased interest in finding better ways to evaluate the language proficiency of young students. The Center for Applied Linguistics (CAL) has been involved in a variety of test development efforts over the years, ranging from simulated oral proficiency interviews for adults to oral proficiency assessments for young children (see Thompson [1995] for a listing of a range of language assessment instruments for children). This paper will describe an alternative assessment instrument that CAL developed to assess oral language skills of students in Spanish immersion programs in the United States. In addition to describing the instrument and rating scale, test results from an immersion program will be presented, and plans for collecting reliability and validity data will be discussed.

II. Description of Test

The purpose of the Student Oral Proficiency Assessment (SOPA) is to determine immersion students' oral proficiency and listening comprehension in a foreign language. Designed for children in grades one through four, the SOPA was developed in response to requests from school districts for an alternative language assessment instrument for students in the lower elementary grades. The instrument is based on the CAL Oral Proficiency Exam (COPE), an interactive immersion assessment developed for fifth and sixth graders in response to a need for "an oral interview-type test that would elicit normal speech and would yield global ratings of proficiency" (Rhodes and Thompson, 1990).

The SOPA was first used in 1991 to evaluate the Spanish partial immersion program at Woodland Elementary School in Oak Ridge, Tennessee, and has been used at various schools since then, e.g., Arlington (Virginia) Public Schools; Foreign Language Immersion and Cultural Studies School, Detroit (Michigan); and Alexandria (Virginia) Public Schools. The SOPA is now being used for oral language assessment of six-to-nine-year-olds in a variety of types of immersion programs, including partial and total immersion and two-way immersion. Recently, the instrument was adapted for use in non-immersion French, German, Japanese, and Spanish elementary school language programs, and a research study, discussed at the end of this paper, is currently underway to evaluate the reliability and validity of the instrument.

The SOPA consists of four parts that are set in an interview format: listening comprehension, informal questions, science and language usage, and story telling. Two students are assessed at a time by two examiners in a non-stressful, friendly environment. The interview takes approximately 10-15 minutes to complete. The goal of the assessment is to show what the students *can* do with language, not what they cannot do. The test aims to get the students to use as much language as possible in a short period so that there will be a large body of data on which to base the ratings. The rating and interviewing tasks are divided between two examiners: one rater and one interviewer. This ensures that the

interviewer can focus entirely on guiding the students to their highest possible level of performance in both listening comprehension and oral fluency, and the rater can focus on rating the students objectively and accurately. The SOPA is conducted entirely in the target language. Ideally, the SOPA should not be used as the only assessment of a student's progress in proficiency development, but should be used in conjunction with teacher observations and other evaluations of the student's daily oral and written work.

III. Rating Scale

Students' language is rated holistically. The SOPA rating scale (see Appendix) uses the first six levels of a nine-level scale from the COPE test, which is based on the proficiency guidelines of the American Council on the Teaching of Foreign Languages. SOPA students receive one of six ratings for comprehension and fluency (whereas the COPE ratings for fifth and sixth graders include comprehension, fluency, vocabulary, and grammar).

The six levels of the rating scale are Junior Novice-Low, Junior Novice-Mid, Junior Novice-High, Junior Intermediate-Low, Junior Intermediate-Mid, and Junior Intermediate-High. The comprehension ratings range from "recognizes a few familiar questions and commands" (Junior Novice Low) to "usually understands speech at normal speed, though some slow-downs are necessary; can request clarification verbally" (Junior Intermediate High). The fluency ratings range from "conversations are limited to an exchange of memorized sentences or phrases" (Junior Novice Low) to "maintains conversation with remarkable fluency but performance may be uneven; uses language creatively to initiate and sustain talk" (Junior Intermediate High). When the SOPA is given annually, a student's ratings are expected to increase gradually, revealing his or her progress in the foreign language.

IV. Test Administration/Components of Test

A quiet room, set up especially for the interview, is ideal for the assessment procedures. The purpose of this arrangement is to create a tranquil, non-threatening environment in which the students can enjoy the activities and feel at ease so that they will be able to speak and listen to their fullest capacity without distractions. Students are evaluated in pairs to facilitate dialogue between each other and between them and the examiners. Because the SOPA is designed to elicit and measure fluency as well as comprehension, the students are encouraged to interact with each other during all the assessment activities.

The interview is tape recorded for later verification of scoring by the raters. During the interview, one examiner serves as the main interviewer, while the other examiner serves as the primary rater, taking notes on both students' language as the interview progresses. The rater marks the students' scores in comprehension and fluency on the rating scale, and makes additional comments about their language skills on the bottom of each sheet. After the interview, the rater and the interviewer compare notes and come to a consensus about the students' ratings. If they need further discussion to agree on a particular student's rating, or just want to "fine tune" the scoring, the tape of the interview is listened to.

The SOPA is comprised of four tasks designed for assessing the various levels of listening comprehension and speaking fluency: listening comprehension, informal questions, science and language usage, and story telling.

Listening comprehension. As the two students enter the room, the interviewers make them comfortable by greeting them and asking them their names, which are put on name tags that they wear. The first part of the test, used as a warm-up, focuses on the students' listening skills. In order to put the students at ease when they first come in, they are handed a bag of plastic fruit and are asked, in Spanish, to empty out the bag and line up the fruit on the table. Focusing only on their listening skills, the interviewer asks the students to point to *el limón, el plátano, las uvas, la fresa, la manzana,* y *la pera.* After the interviewer assesses the

students' comprehension, she or he then asks questions dealing with the color of the fruit (*amarillo, morada, roja, y verde*), the number of each type, and the students' favorites in order to elicit responses for assessing fluency. Using the fruit as manipulatives, students are then asked to respond to various commands, such as *Pon el plátano encima de tu cabeza* (Put the banana on your head), *Pon las uvas debajo de la mesa* (Put the grapes under the table), and *Pon la manzana dentro del libro* (Put the apple inside the book). Still using the fruits, the students are asked to perform a cognitively more complex task, such as naming some fruits and colors that are not represented by the fruits on the table.

Informal questions. After the fruit activity, the interviewer asks the students a few personal questions in order to further assess fluency and comprehension for basic language concepts. Some examples of these questions are *¿Cuántos años tienes?* (How old are you?) *¿Cuántos hermanos tienes?* (How many brothers and sisters do you have?) and *¿Tienes un animal en tu casa?* (Do you have a pet at home?).

Science and language usage. At this point, the students usually feel more at ease with the procedure, so the interviewer begins the third part of the assessment which is to review their language skills in science. Knowledge of science concepts and language used to talk about science are measured by a series of four pictures that show: (1) a father and little girl planting a small tree, (2) the little girl watering the small plant, (3) the plant growing in the sunshine, and (4) a full-grown tree. The children are told that these are a series of pictures, in order, and are asked to describe what is going on in each picture. The first picture is prompted by the question, *¿Qué están haciendo el papá y la niña?* (What are the father and daughter doing?) If the students don't offer any description for the second picture, they are asked, *¿Que hace la niña?* (What is the girl doing?) For the third picture they are asked, *¿Qué está pasando aquí?* (What's going on here?) And for the final picture, *¿Qué es? ¿Es grande o pequeño?* (What is this? Is it big or small?) If the students don't spontaneously produce language about the pictures, the examiner prompts with specific questions and the students are asked to identify objects and people.

Story telling. The final part of the assessment is the story telling. The students are handed a book with a story that they have either heard or read, in either language, and are asked to tell the story in Spanish by describing what is happening in the pictures. The interviewer goes through the book page by page with the students, prompting them with questions on each page if they don't initiate anything on their own. The story *Goldilocks and the Three Bears* is the story most often used, and the following questions can be asked: *¿Quiénes son?* (Who are they?) (pointing to the three bears and Goldilocks). *¿Dónde viven?* (Where do they live?) (pointing to the house). *¿Qué hace el bebé?* (What is the baby doing?) (pointing to the baby bear eating porridge). *¿Quién está entrando en la casa?* (Who is entering the house?) (pointing to Goldilocks entering the house). *¿Qué hace la niña?* (What is the girl doing?) (pointing to Goldilocks eating porridge). *¿Qué son estos? ¿Cuál es grande? ¿Mediana? ¿Pequeña?* (What are these? Which is big? medium? small?) (pointing to chairs). *¿De quién es esta silla?* (Who's chair is this?) pointing to papa bear's chair, mama bear's chair, baby bear's chair). *¿Qué le pasó a Goldilocks?* (What happened to Goldilocks?) (pointing to her sitting in the chair and breaking it). *¿Qué está haciendo Goldilocks? ¿De quién es la cama?* (What is Goldilocks doing? Who's bed is this?) (pointing to her in baby bear's bed). *¿Cómo están los osos? ¿Están enojados o contentos? ¿El bebe está triste o feliz?* (How are the bears? Are they mad or happy? Is the baby sad or happy?) (pointing to their return to their messy house). *Al fin, ¿qué le pasó a Goldilocks?* (Finally, what happened to Goldilocks?) (pointing to her running away).

Equipment/supplies needed. The following supplied are needed to administer the SOPA: pieces of fruit (plastic or rubber eraser type); picture sequence of science concepts; storybook with pictures, such as "Goldilocks and the Three Bears" (cover up all text), the SOPA rating scale; name tags for the students; cassette tape recorder; and cassette tapes.

V. Results

To provide a sense of how the SOPA has been used and the scores that have been obtained, the results of test administration in a Spanish two-way immersion program will be described. The program, located in a suburban Virginia elementary school that includes grades 1 - 5, teaches approximately half the daily curriculum in Spanish, the other half in English. The program is considered "two-way" because the student body is made up of both Spanish language background students and native English language background students. Researcher Beverly Boyson assessed twenty-nine second graders in the immersion program: fourteen students had native Spanish language backgrounds, and fifteen had native English and/or other language backgrounds.

Overall results. The second grade students performed well in the preliminary listening comprehension exercise. Nearly all of the students were able to identify the fruits and their colors, responding nonverbally to questions such as *¿Cuál es la manzana?* (Which one is the apple?) and *¿Cuál fruta es amarilla?* (Which fruit is yellow?). Also, most students could respond verbally to questions such as *¿Cómo se llama esta fruta?* (What is this fruit called?), *¿De qué color es esta fruta?* (What color is this fruit?) and *¿Cuál es tu fruta favorita?* (What is your favorite fruit?).

While some students hesitated when asked to react to the commands that required recognition of prepositions (e.g., *encima de, debajo de, dentro de*), most students were able to understand the commands. A few of the less advanced students from English language backgrounds required prompting during this initial exercise, while the more advanced students, from both Spanish and English language backgrounds, reacted quickly to the commands and often initiated talk about the fruit.

Most students were able to understand basic informal questions, such as *¿Cuántos años tienes?* (How old are you?) and *¿Tienes hermanos?* (Do you have brothers/sisters?). When students appeared to comprehend the initial questions with ease, they were asked more complex questions that required them to elaborate on other topic areas. Less

advanced students sometimes reverted to English either completely or partially in their responses, while more advanced students rarely reverted to English and were sometimes able to extend their discourse beyond the expected level of competence for the task.

Nearly all students were able to identify the main objects in the sequence of pictures for the science exercise. Students in the higher levels from both Spanish and English language backgrounds were able to understand all of the exercise. Many of these students could produce sentence-level speech. While some of the English language background students were unable to form complex sentences that were grammatically accurate, the majority of these higher level students were able to produce simple descriptive sentences sequentially. Those with the highest levels of fluency used specific vocabulary and complex structures. The following is an interaction between two seven-year-old native-English-speaking second graders (in their second year of Spanish immersion) and the interviewer ("maestra") on the science portion of the SOPA.

Transcript of Second Graders During Science Portion of SOPA Interview

---

Maestra:	*Ahora quiero enseñar unos dibujos y quiero que me digan todo lo que está pasando aquí en los dibujos. ¿Qué está pasando aquí?*
Kathy:	*Planta un semilla.*
Maestra:	*¿Quienes?*
Kathy:	*La niña y el padre.*
Maestra:	*Sí, muy bien. ¿Y que está pasando aquí, Kathy?*
Kathy:	*La niña es pon agua en la planta.*
Maestra:	*Sí, está poniendo mucha agua. ¿Y aquí, Victoria?*
Victoria:	*En la planta está creseando a un árbol porque la sol está brillando a la planta.*
Maestra:	*Sí, muy bien. ¿Cómo se llama esta parte de la planta?*
Victoria:	*Los...es un dibujo.*
Maestra:	*Sí.*
Victoria:	*¿Y las frutas?*
Maestra:	*Las raices. ¿Y ésta? ¿Cómo se llama, Kathy?*
Kathy:	. . . (whispers in English: *I forget what it's called.*)
Maestra:	*Es una letra silenciosa al principio.*
Kathy:	*¡Hoja!*
Maestra:	*Sí, muy bien. ¿Qué necesitan las plantas para crecer?*
Victoria:	*Necesita el sol, la aire, la lluvia, y la tierra.*
Maestra:	*La tierra, cierto. Muy bien. ¿Y qué pasó al final, Kathy?*
Kathy:	*La planta es un árbol.*
Maestra:	*Es cierto.*

As can be seen in the example, the children understood almost everything that was asked of them about the picture series, and were able to describe the growth of a plant, offering the reasons why a plant grows. Since the goal of the interview is to obtain as large a language sample as possible to show what the students can do, it is important to highlight all the aspects of the interview that showed what the students can do, both linguistically and scientifically. For example, the students identified the people, described the girl watering the plant, described that the plant was growing because of the sunlight, and named four things that plants need to grow: sun, air, rain, and soil. Grammatically speaking, if this exchange were examined in a more traditional error correction mode, it would, of course have to be pointed out that the children used incorrect adjective agreement (*un semilla* instead of *una*); wrong verb forms (*la niña es pon agua* instead of *pone* or *está poniendo;* and incorrect pronunciation (*creseando* instead of *creciendo*); among other things. Since the student's ability to communicate is the overall concern of the SOPA, grammar only is an issue if it interferes with that ability.

As expected, Spanish language background students were able to express the sequence more elaborately, using varied sentence patterns including the subjunctive, e.g., *Aquí están plantando una semilla para que crezca* (Here they are planting a seed so that it will grow).

In the final phase of the assessment, the students were asked to create a storyline for Goldilocks and the Three Bears. Nearly all students, being familiar with the story, made attempts to explain it in Spanish. Many of the Spanish language background students were able to produce full sentences and maintain simple narratives to describe the story, often using past and progressive tenses, e.g., *Se quebró la silla* (The chair broke) and *Está comiendo la avena* (He is eating the cereal). Many higher-level Spanish language background students were able to accurately use subjunctive, e.g., *La mama está poniendo, pues para que se seque las ropa . . .* (the mother is putting [clothes on the line], so that the clothes dry . . .), and present perfect tense, e.g., *Alguien se ha sentado en mi sillas*

(Someone has sat in my chair). Some students were able to distinguish the usage of simple past and the imperfect, e.g., *Alguien probó mi sopa* (Someone tasted my soup) and *Estaba cansada y quería dormir* (She was tired and wanted to sleep). Others included creative dialogue between the characters or added details to the story such as the feelings of the characters and descriptions of the scenes.

A few English language background students were able to produce complete sentences and react spontaneously to normal conversation during the storytelling segment. Like their Spanish background counterparts, these higher level students were able to initiate talk and produce original responses. In general, lower-level English language background students were able to identify objects and/or characters in the story with some prompting and could understand basic questions which they responded to both verbally and non-verbally, distinguishing between true/false statements. Students from both language groups who had difficulty explaining an action were asked to identify as much as they could in a particular scene and were sometimes prompted towards an appropriate response.

Comprehension and fluency ratings. Both Spanish language background students and English language background students were rated on a scale from Junior Novice-Low (number 1 on the SOPA scale) to Junior Intermediate-High (number 6 on the SOPA scale) on comprehension and fluency (see Figures 1 and 2).

Figure 1. Grouping of Students According to Comprehension Skills on the SOPA Rating Scale for Spanish:

1	2	3	4	5	6
Jr. Nov. Low	Jr. Nov. Mid	Jr. Nov. High	Jr. Inter. Low	Jr. Inter. Mid	Jr. Inter. High
eee	ee	eee		eee	eeee sssssssssssss
10%	8%	10%	0%	0%	62%

e = one English language background student
s = one Spanish language background student
Total = 29 students
Note. Percentages are approximate

Figure 2: Grouping of Students According to Fluency on the SOPA Rating Scale for Spanish:

1	2	3	4	5	6
Jr. Nov. Low	Jr. Nov. Mid	Jr. Nov. High	Jr. Inter. Low	Jr. Inter. Mid	Jr. Inter. High
eeeee	eeee s	eee	e	e ssss	e sssssssss
17%	17%	10%	3%	17%	34%

e = one English language background student
s = one Spanish language background student
Total = 29 students
Note: Percentages are approximate

All of the Spanish language background students scored at the highest level, Jr. Intermediate-High, for comprehension. The majority of these students demonstrated Jr. Intermediate-Mid and Jr. Intermediate-High fluency levels. The English language background students had a broad range of comprehension and fluency levels, scoring between Jr. Novice-Low and Jr. Intermediate-High. Half scored in the Jr. Intermediate and half in the Jr. Novice levels for comprehension, while the majority scored in the Jr. Novice levels for fluency. As is expected of second graders in a partial immersion program, their comprehension skills were as strong or stronger than their speaking skills.

The results indicate that a greater number of Spanish language background students were at the upper end of the scale. A larger number of English language background students were at the lower end of the scale; however, a few of these students placed in the Jr. Intermediate levels for fluency, exceeding expectations. In general, the distribution of ratings for the students reflected the expected levels of proficiency for second graders in a partial immersion program.

Summary. The results of the second graders' performance on the SOPA are positive. Both Spanish language background students and English language background students exhibited an impressive range of comprehension and fluency levels during the various listening and speaking activities. The majority of Spanish language background students were rated in the Jr. Intermediate categories, indicating their ease with the language and their ability to satisfy most academic and social functions. In addition, they

were often able to participate in full discussion on familiar topics while a few were also able to expand on unfamiliar topics.

The English background speakers demonstrated an ability to comprehend and participate in most of the assessment tasks, with almost half of the students rated at the Jr. Intermediate levels in comprehension. Their growing proficiency was evident in their ability to identify and describe various objects and carry out commands. Some students showed definite signs of mastering a second language as they attempted to engage in more creative dialogue during the exercises that demanded greater production. These students spoke creatively at the sentence-level and initiated some talk.

The oral proficiency of these second graders indicates the valuable instruction they are receiving in the Spanish language. The majority of the Spanish language background students have been able to maintain and develop an impressive level of proficiency in their native language. The English language background students have demonstrated an impressive range of proficiency for children who are learning a second language, considering that their exposure to Spanish is often limited to the classroom setting.

VI. Future Research: Validity and Reliability Research Study

The Center for Applied Linguistics, in conjunction with the Iowa State University National K-12 Foreign Language Resource Center, is now in the process of revising and validating the SOPA. Although it is based on a reliable test that has been validated, the SOPA itself has never been validated and has not been formally packaged for distribution to teachers. The validation of an instrument such as this is critical if the United States is going to continue its pursuit of excellence in early language programs. As more and more programs are created and continue to focus on accountability and standards, the profession needs to be able to demonstrate how well students are doing. In order to do this in a field that has few assessment methods for children's language, it is important to develop valid and reliable instruments and then train teachers in the administration and interpretation of these tests.

This small-scale research study includes a revision of the rating scale, the development of a SOPA version for regular (non-immersion) language programs, clinical testing, validity and reliability testing, and the development of a test administrator's manual. Validity testing will be conducted with students in three kinds of programs: immersion, content-based, and regular (non-immersion) language programs. At least 50 students at each site will be tested in grades 1 - 4, with a focus on students who have had at least 2 years of language. Sites will be selected from schools of the teachers who receive SOPA administration and rater training at a 1997 summer assessment institute. The IDEA Proficiency Test will also be administered to each of the students at the sites to test the validity of the instrument. Additional data on student achievement will be gathered at each school to be used in evaluating the students' performance on the SOPA. If resources allow, and if additional teachers volunteer, additional programs of each type will be included in the validity testing.

The project activities are the following:

(1) revise SOPA rating scale using newly developed national foreign language standards and immersion benchmarks,

(2) develop alternate form of SOPA for regular (non-immersion) language students,

(3) conduct SOPA clinical testing (immersion and non-immersion versions); revise instruments and scales as needed,

(4) administer SOPA and comparable instrument at three sites (immersion, content-based, and non-immersion) for validity purposes,

(5) collect validity and reliability information on the instrument,

(6) develop administrator's manual.

The following is a detailed description of the six activities:

Rating scale revision. The first task, already completed, was to revise the SOPA rating scale, using input from classroom teachers, the national foreign language standards, and the newly-developed immersion benchmarks from Arlington, Virginia Public Schools. The comment most often received from teachers in the past was that the scale was useful but needed some adjustment so that it more accurately reflected the language of students in content-based and immersion programs. It is hoped that the revision has fine tuned the scale by clarifying the

descriptions of each level and, where appropriate, adjusting the language to reflect the content of the language program.

Alternate SOPA. The next step was to adapt the SOPA for students involved in regular language programs (those that meet from one to five times a week for less than 50 minutes/day and focus on language *per se*, with little academic subject matter included). The rationale for adapting the SOPA for regular language programs is that they make up the majority of early language programs in the United States and teachers are constantly requesting more accurate assessment measures for their students. Currently there are few, if any, standardized oral language assessments in use in regular elementary school programs.

The SOPA adaptation for non-immersion students includes two major changes. Since science is not taught in the regular language class, the *science and language usage* section was exchanged for an interactive "peel and stick" dollhouse activity, where the examiner assesses comprehension by asking the students to place certain plastic colorform objects and people in different rooms of the house. Second, the *story telling* section was replaced with a more appropriate descriptive activity involving a picture of a classroom. The *listening comprehension* and *informal questions* section are appropriate for regular language students and remain the same. The important question of whether regular language students can be rated fairly on the same scale as immersion and content-based students is still being addressed. Tentatively, through discussions with teachers and specialists in the field, it has been agreed that the revised SOPA scale is appropriate for both immersion and regular language students.

Clinical testing. The clinical testing of the alternate SOPA and the revised rating scale will take place with a few students in regular and immersion programs in the Washington, DC area. The students will be administered the alternate SOPA or the immersion SOPA to see if the tests are "child-friendly" and to review the accuracy of the rating scales. An additional purpose of the clinical testing is to get feedback from the students on the content of the test (Was the subject matter appropriate? Had they learned the concepts already?), the test administration (Could they follow the directions? Did they understand what they were supposed to do?), and their overall

impressions of the test (Did they enjoy participating in the test? What did they learn from it?). After observing the children participate in the assessment, discussing their views on the test, and evaluating the adequacy of the rating scale, appropriate revisions will be made.

Validity testing: SOPA and IDEA Proficiency Test. The validity testing of the alternate SOPA and immersion SOPA will take place at three sites: one immersion, one content-based, and one regular language program. Participants at the summer assessment institute will receive training in SOPA administration and rating. Testing sites will be selected from participants, representing the range of programs, who volunteer to have their schools participate. These teachers will serve as local site coordinators and will: (1) administer the SOPA to the designated students in their school; (2) collect background data on the students from the classroom teacher including other test scores, Student Oral Proficiency Rating scores, grades in language class, and other relevant information; and (3) coordinate the IDEA Proficiency Test administration. The local site coordinator will also "debrief" the students after the testing to get feedback on the test administration and any comments on the content of the test.

The IDEA Proficiency Test, a previously validated instrument, was selected as the oral proficiency test to be administered at the same time as the SOPA to assess its concurrent validity. Among the many types of test validity, concurrent validity, or the extent to which a test score corroborates the result of an independent external criterion measure administered at the same point in time, will be used. For reliability purposes, a different rater will be used for the SOPA and the IDEA Proficiency Test. This way, artificially inflating the validity index will be avoided -- a possible result of using the same rater who might assign a second rating based on knowledge of the student's performance on the previous test. The administration of the test will take from 5-15 minutes, depending on the proficiency level of the student.

The IDEA Proficiency Test was designed to measure native Spanish speakers' oral proficiency in Spanish. The test consists of 83 items, with each item testing one of six oral language skill areas: syntax, morphology, lexicon, phonology, comprehension, and oral expression. The student is required to respond to the questions presented either verbally or

visually. Student performance is rated on a scale from A-F, with an additional possible category of M, which designates mastery of the test. The scale can then be collapsed into a three-category scale: NSS (Non-Spanish Speaking), LSS (Limited Spanish Speaking), and FSS (Fluent Spanish Speaking).

Validity and reliability. Validity refers to the extent to which a test measures what it is intended to measure. In addition to the concurrent validity measure mentioned above, student scores on another instrument, the Student Oral Proficiency Rating, will also be collected at the sites as an additional criterion measure.

Operationally, the concurrent validity index of the SOPA will be measured by calculating the Pearson product-moment correlation between the total SOPA and the IDEA Proficiency Test. The total SOPA will be coded as the sum of the two subscores, comprehension and fluency. Each subscore ranges from 1 (Junior Novice Low) to 6 (Junior Intermediate High). The IDEA Proficiency Test will be coded on a scale of 1-7, representing the original A-F scale plus M.

Additionally, content validity will be measured by experts in the field who will review it for face validity, i.e., whether the content is appropriate and whether they would "certify" it to measure what it is supposed to measure.

Inter-rater reliability will be assured by the intensive training that the test administrators will receive at the 1997 summer institute. During the training, teachers will practice rating sample tapes and their scores will be compared with the correct scores. Teachers will continue practicing until most of their scores are the same as a master rater's scores and other scores are no more than one level off). Test-retest reliability (when the same test is administered again to assess whether students perform at approximately the same level each time they take the test) will be an optional reliability test that will be conducted with a subsample of students if time allows.

Administrator's manual. The administrator's manual will include background information on the SOPA, research results of the clinical testing, the rating scale, a discussion of the rating system, audio tapes and transcripts of sample interviews, and a guide to the rating

process. The manual will be suitable for the training of all versions and all languages of the SOPA.

VII. Conclusion

Results of SOPA administration with students in various language programs during the last six years have been quite positive. Test administrators have found that a large language sample can be gathered during the various tasks of the interview; a wide range of language skills, both academic and social, can be assessed; and, most importantly, interviewers and students view the interview as a positive learning experience. Test administrators have reported that, for the most part, they are able to elicit a language sample that accurately reflects the students' everyday language.

The one area that needs additional attention for future administrations of the SOPA is the training of the raters. As is common with other global rating scales, it is challenge to develop a cadre of raters whose scores are reliable and consistent. Although ample time is allowed for rater training and assurance of inter-rater reliability at the 1997 summer institute, it remains a critical issue to address for future rater training. Practically speaking, elementary school teachers who are trained as interviewers and raters are from schools across the country, and it is often difficult to follow up with the necessary additional training to ensure that the raters are consistent with their ratings. In the months ahead, the SOPA development team will pursue various options in rater training, given the parameters of our pool of raters, in an attempt to provide for more consistent rating of students language over time.

References

Rhodes, N. & Thompson, L. (1990). "An Oral Assessment Instrument for Immersion Students: COPE." A.M. Padilla, H.H. Fairchild, & C.M. Valadez (eds.), Foreign Language Education: Issues and Strategies (p. 75-94). Newbury Park, CA: Sage.

Thompson, L. (1995). K-8 Foreign Language Assessment: A Bibliography. Washington, DC: Center for Applied Linguistics.

Appendix

Rating Scale for Student Oral Proficiency Assessment (SOPA)

School _____ Grade ___ City/State _____ Examiners _____ Date _____

	JR. NOVICE LOW	JR. NOVICE MID	JR. NOVICE HIGH	JR. INTERMEDIATE LOW	JR. INTERMEDIATE MID	JR. INTERMEDIATE HIGH
Compre-hension	Recognizes isolated words and high-frequency utterances.	Understands predictable questions, statements and commands in specified topic areas (with strong contextual support), though at slower than normal rate of speech or with repetitions.	Understands simple questions, statements and commands in specific topic areas, and some new sentences (with strong contextual support). May require repetition, rephrasing or slower rate of speech.	Understands new sentence-level questions and commands in a limited number of content areas (with strong contextual support). Follows conversation at a fairly normal rate of speech.	Understands sentence-level utterances in new contexts though some slow-downs may be necessary. Carries out commands without prompting. May show some difficulty on unfamiliar topics.	Usually understands speech at normal speed. Compre- hension problems seldom evident on everyday topics. Understands longer stretches of connected discourse on a number of topics. Can request clarification verbally.
Fluency	Oral production consists of isolated words or high-frequency phrases.	Uses a limited number of words and phrases within predictable topic areas. Long pauses are common. May attempt sentences in the target language, but frequently completes them with gestures or use of native language.	Uses high frequency utterances with reasonable ease. Signs of originality and spontaneity begin to emerge. Attempts to create sentences with some success, but is unable to sustain sentence-level discourse.	Able to handle a limited number of everyday social and academic interactions. Maintains simple conversations at the sentence level by creating with the language, although in a restrictive and reactive manner with linguistic inaccuracy.	Shows evidence of spontaneity in conversation. Sometimes initiates talk without relying on questions or prompts. Gives simple descriptions successfully, although inaccuracies are present.	Maintains conversation with increasing fluency, but errors are still present and vocabulary is limited. Uses language creatively to initiate and sustain talk. There is emerging evidence of connected discourse in simple narratives.

Student's Name:

Examiners' Comments:

This scale was developed by the Center for Applied Linguistics (and adapted from the CAL Oral Proficiency Exam), 1996